

Consortium 2018 State-by-State Comparisons

D. J. McRae, Ph.D.

October 11, 2018

[Updated Late November, 2018]

When two consortiums of states were chosen by the US Department of Education in 2010 to develop new statewide assessment systems, one of the purposes was to generate state-by-state comparable achievement data. Roughly 45 states initially signed up for potential use of consortium tests, but by 2015 when the new tests were ready for their first “operational” use, only 18 states administered the Smarter Balanced tests and 11 states (plus the District of Columbia) administered the PARCC tests, representing just under 50 percent of total K-12 enrollments across the country. In 2018, 11 states administered Smarter Balanced tests, and 4 states (plus DC) administered PARCC tests, representing 29 percent of K-12 enrollments across the country. In addition, Louisiana, Massachusetts, Colorado, and Rhode Island (PARCC) and Michigan (Smarter Balanced) used public domain test questions for their statewide tests but did not use full consortium protocols and hence are not included in this set of state-by-state comparisons. A footnote on page 5 notes this will be the last year that comparable information will be available for PARCC states.

The data charts on pages 2 and 4 provide state-by-state results for the Smarter Balanced and PARCC states, respectively, for spring 2018 testing. The results are expressed as “percents meeting target” grade-by-grade for English Language Arts and Mathematics, along with average percents across grades. On pages 3 and 5, the average gain scores for each state for 2015 thru 2018 are provided, respectively, for Smarter Balanced and PARCC states (plus DC). Notes describing the data in the charts are provided at the bottom of pages 3 and 5, respectively, for SBAC and PARCC. Results represent preliminary data released by states in many cases, with final data to be released later. Annual percent-meeting-target results from Smarter Balanced states are not comparable to results from PARCC states, and even within consortia there may be some differences in test administration or reporting practices across states. However, within consortiums, the comparability of scores is sufficient for general comparisons.

Finally, it is fair game to average gain scores for ELA and Math for each state to produce an annual overall gain score for 2016, 2017, and 2018 results. In the early 2000’s, highly respected educational measurement expert Bob Linn testified before Congress that 3- to 4-percentage point annual gains for statewide testing programs could be characterized as good to very good, and 2-point annual gains were typical. With this background, it is fair game to interpret each annual gain score as a letter grade based on a 4.0 grade point average (GPA) metric, with 4.0 being an A, 3.0 being a B, 2.0 being a C, 1.0 being a D, and 0.0 being an F. Annual gain scores for 2016 thru 2018 for all consortium states (plus DC) are provided on page 6. The

year-to-year gain scores and assigned letter grades on page 6 are comparable across consortiums.

Smarter Balanced 2018 State-by-State Comparisons [Level 3 & Above Percents]

Compiled by D. J. McRae, Ph.D.

			<u>English/Language Arts</u>						
	<u>Grade</u>		<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	<u>Ave</u>
1	California	48	49	49	48	50	49		48.8
2	Connecticut		53	55	58	54	55	56	55.2
3	Delaware	52	55	58	52	54	53		54.0
4	Hawaii	53	51	56	53	52	55		53.3
5	Idaho	50	50	55	53	54	54		52.7
6	Montana	50	49	53	51	51	49		50.5
7	Nevada	46	48	50	44	47	47		47.0
8	Oregon	47	50	55	52	57	56		52.8
9	South Dakota	51	50	53	52	56	56		53.0
10	Vermont	50	53	55	53	57	57		54.2
11	Washington		56	57	59	56	60	59	57.8
	Averages	51	52	55	52	54	54		52.7

			<u>Mathematics</u>						
1	California		49	43	36	38	37	37	40.0
2	Connecticut	54	51	45	44	44	43		46.8
3	Delaware		54	50	43	40	39	39	44.2
4	Hawaii	55	47	43	42	37	38		43.7
5	Idaho	52	48	43	44	44	41		45.3
6	Montana	49	45	40	39	39	37		41.5
7	Nevada	48	42	36	32	31	30		36.5
8	Oregon	46	43	40	38	42	41		41.7
9	South Dakota	55	49	41	44	48	48		47.5
10	Vermont	52	49	42	41	44	42		45.0
11	Washington	58	54	49	48	49	48		51.0
	Averages	52	47	42	41	41	40		43.9

Smarter Balanced 2015-18 Gain Scores

		<u>English/Language Arts</u>						
		2015	2016	2017	2018	15-16	16-17	
17-18	<u>State</u> Gain		ELA	ELA	ELA	Gain	Gain	
1	California +2.0	42.3	46.7	46.8	48.8	+4.4	+0.1	
2	Connecticut +1.0	xx*	55.8	54.2	55.2	xx*	-1.6	
3	Delaware +0.2	51.7	54.8	53.8	54.0	+3.1	-1.0	
4	Hawaii +4.1		47.7	50.5	49.2	53.3	+2.8	-1.3
5	Idaho +1.5		49.7	51.8	51.2	52.7	+2.1	-0.6
6	Montana +0.7	xx*	50.0	49.8	50.5	xx*	-0.2	
7	Nevada +0.8	xx*	48.3	46.2	47.0	xx*	-2.1	
8	Oregon +1.3		53.8	53.3	51.5	52.8	-0.5	-1.8
9	South Dakota +3.2		47.5	51.2	49.8	53.0	+3.7	-1.4
10	Vermont	53.7	56.5	52.5	54.4	+2.8	-4.0	+1.9
11	Washington +0.8	55.5	57.8	57.0	57.8	+2.2	-0.8	

		<u>Mathematics</u>							
		2015	2016	2017	2018			15-16	16-17
		17-18							
Gain	<u>State</u>	Gain	Math	Math	Math	Math			Gain
+1.8	1	California	34.2	37.3	38.2	40.0		+3.3	+0.9
	2	Connecticut	40.3	44.2	45.8	46.8		+3.9	+1.6
	3	Delaware	40.7	43.7	44.5	44.2		+3.0	+0.8
		+1.0							
		-0.3							
+0.2	4	Hawaii		42.2	43.0	43.2	43.7		+0.8
		+0.5							

0.0	5	Idaho	40.8	43.3	43.3	45.3	+3.5	
	+2.0							
+0.3	6	Montana	xx*	41.0	41.2	41.5	xx*	+0.2
	7	Nevada		xx*	33.8	33.3	36.5	xx* -0.5
	+3.2							
	8	Oregon		43.5	42.8	41.8	41.7	-1.2 -1.0
	-0.1							
+1.3	9	South Dakota		41.2	44.5	45.8	47.5	+3.3
	+1.7							
+0.8	10	Vermont	43.2	46.7	44.2	45.0	+3.4	-2.5
-0.2	11	Washington	49.8	51.5	51.2	51.0	+1.7	-0.3

Notes for Smarter Balanced Data:

All averages and gains are based on Grade 3-8 data. Only selected states use Smarter Balanced HS tests.

Montana and Nevada participated in Smarter Balanced testing in 2015, but both states experienced technology difficulties that prevented generation of representative scores for the entire state. This circumstance prevents calculation of selected gain scores. Connecticut discontinued the Performance Task for the ELA test in 2016, so for comparability reasons the 15-16 ELA gain score is not recorded.

PARCC 2018 State-by-State Comparisons [Level 4 & Above Percents]

Compiled by D. J. McRae, Ph.D.

		<u>English Language Arts</u>						<u>Ave</u>
<u>Grade</u>		<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	
1. Dist Columbia		31	35	34	31	39	33	33.8
2. Illinois	37	39	36	34	40	36		37.0
3. Maryland	39	43	42	39	46	41		41.7
4. New Jersey	52	58	58	56	63	60		57.8
5. New Mexico	29	29	31	28	29	29		29.2
Averages	38	41	40	38	43	40		39.9

		<u>Mathematics</u>						<u>Ave</u>
<u>Grade</u>		<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>	<u>8</u>	
1. Dist Columbia		41	34	33	24	25	xx	31.4
2. Illinois	38	32	31	27	31	31		31.7
3. Maryland	42	39	38	32	xx	xx		37.8
4. New Jersey	53	49	49	44	43	xx		47.6
5. New Mexico	32	26	28	21	20	xx		25.4
Averages	41	36	36	30	30	xx		34.8

PARCC 2015-18 Gain Scores

		<u>English/Language Arts</u>					
		2015	2016	2017	2018	15-16	16-17
17-18	<u>State</u>		Ave	Ave	Ave	Gain	Gain
Gain							
1. Dist Columbia	+6.0	+1.3	26.2	26.5	32.5	33.8	+0.3
2. Illinois	0.0	36.0	36.5	37.0	37.0	+0.5	+0.5
3. Maryland	+1.2	38.8	38.7	40.5	41.7	-0.1	+1.8
4. New Jersey	+1.8	50.0	52.8	56.0	57.8	+2.8	+3.2
5. New Mexico	+2.5	23.2	24.5	26.7	29.2	+1.3	+1.2

		<u>Mathematics</u>					
1. Dist Columbia	+0.7 +1.2	25.3	29.5	30.2	31.4	+4.2	
2. Illinois	+0.5	28.8	31.3	31.2	31.7	+2.5	-0.1
3. Maryland	+0.8	31.8	37.3	37.0	37.8	+5.5	-0.3
4. New Jersey	+1.8	41.0	45.6	45.8	47.6	+4.6	+0.2
5. New Mexico	+3.0	21.4	23.0	22.4	25.4	+1.6	-0.6

Notes for PARCC Data:

Most PARCC states utilized PARCC End-of-Course Math tests for the High School level rather than PARCC grade level tests. Since course taking patterns differ from state-to-state, HS results are not included.

All averages and gains reflect grades 3-8 only, and the xx under a grade means the state does not uniformly administer PARCC grade level tests to all students at that grade level. All xx's reflect administrations of Algebra and/or Geometry End-of-

Course tests to grade 7 and 8 students taking these courses, rather than the regular grade level tests. The pattern of xx's in the chart for 2018 results are identical to the patterns for 2015 thru 2017. Averages are based on grade levels with comparable results.

The PARCC consortium provided item retirement / item replacement / and resulting cut score adjustment services to their member states in the past, but individual states will be responsible for these services in future years. As a result, states using PARCC items as a base for future tests will no longer have comparable gain scores beyond 2018.

ELA / Math Gain Score Averages by Year with GPA's and Letter Grades

		Ave Gain		GPA		Ave Gain		GPA		Ave Gain		GPA	
Letter		15-16		Letter		16-17		Letter		17-18			
SBAC													
1.	California	3.75	A	0.50	D-	1.90	C						
2.	Connecticut	xx	--	0.00	F	1.00	D						
3.	Delaware	3.05	B	-0.10	F	-0.05	F						
4.	Hawaii	1.80	C	-0.55	F	2.30	C+						
5.	Idaho	2.30	C+	-0.30	F	1.75	C						
6.	Montana	xx	--	0.00	F	0.50	D-						
7.	Nevada	xx	--	-1.30	F	2.00	C						
8.	Oregon		-0.60	F	-1.40	F	0.60						
9.	South Dakota		3.50	A-	-0.05	F	2.45	C+					
10.	Vermont	3.65	A-	-3.25	F	1.35	D+						
11.	Washington	2.00	C	-0.55	F	0.30	F						
Averages		2.46	C+	-0.90	F	1.28	D+						
PARCC													
1.	Dist Columbia		2.25	C+		3.35	B+				1.25		
	D+												
2.	Illinois	1.50	C-	0.20	F	0.25	F						
3.	Maryland	2.70	B-	0.75	D	1.00							
	D												
4.	New Jersey	3.70	A-	1.70	C-	1.80	C						
5.	New Mexico	1.45	D+	0.80	D	2.75	B						
Averages		2.40	C+	1.17	D	1.41	D+						

GPA to Letter Conversions:

A = 3.50 to 4.49,
 B = 2.50 to 3.49,
 C = 1.50 to 2.49,
 D = 0.50 to 1.49,
 F = Less than 0.50.

Within each range, the higher range of 0.25 to 0.49 merits a plus sign, the lower range of 0.50 to 0.74 merits a minus sign.

Equal to or greater than 4.50 merits an A++.

Observations for Smarter Balanced and PARCC 2018 State-by-State Comparison Scores

Smarter Balanced vs PARCC Results, ELA vs Math Results, Trends Across Grades, and Gain Results

It is clear students score better on Smarter Balanced tests than on PARCC tests. Smarter Balanced states averaged 53 percent meeting targets for ELA and 44 percent for Math, while PARCC states averaged 40 percent meeting targets for ELA and 35 percent for Math. In 2004, Bob Linn noted differences of 3 to 4 percent are clearly meaningful. Differences of close to 10 percent for ELA are clearly very meaningful differences.

In addition, consortium ELA tests averaged 46.5 percent meeting target, while consortium Math tests averaged 39.5 percent meeting target, another meaningful difference.

While there may be demographic differences between the two cohorts of states, or there may be differences for implementation of common core instruction, it is unlikely either of these reasons would cause the large differences in Smarter Balanced scores vs PARCC scores. Rather, it is likely that the differences between Smarter Balanced and PARCC results are due to the tests themselves, either in the difficulty of the items or in the setting of threshold scores for the respective targets upon which the data in the charts are based. Perhaps the best way to describe the differences between Smarter Balanced and PARCC results is simply that PARCC has the more difficult set of tests.

A look at trends across grades shows no obvious trends for ELA results for both consortiums, but do show declining results for Math as the grades increase for both consortiums. These trends across grades are very similar to the trends across grades found for both Smarter Balanced and PARCC for 2015, 2016, and 2017 results.

The gain scores on pages 3 and 5 as well as the gain scores on page 6 are comparable across Smarter Balanced and PARCC. The annual gains for 2016 through 2018 show both consortiums had somewhat better than typical gains for 2016, considerably lower gains for 2017 [with most Smarter Balanced states showing actual declines], and recovery for less than typical gains for 2018. The pattern of gains for Smarter Balanced show extreme changes from year-to-year, notably different than a more modest pattern for PARCC. One might describe the 2016 thru 2018 gains for Smarter Balanced states as reflective of a Level 5 roller coaster, while the gains for PARCC states reflect a less extreme level 2 roller coaster. The GPA-like metrics for the overall patterns of annual gains translate into Letter grades that communicate these differences very accurately.

Due to the number of student scores entering into these consortium-wide calculations, increases or declines in results of perhaps 0.1 to 0.2 percentage points may be considered “statistically significant.” However, the use of theoretical statistical significance calculations for these analyses of statewide test results is questionable. From a practical perspective, increases or decreases of 0.5 percentage points may be considered “meaningful” changes, and increases/decreases of more than 1.0 percentage points should be considered as “very meaningful” changes, similar to typical interpretation of 4.0 GPAs.

Other Considerations

It should be noted that changes (or lack of needed changes) in the tests between 2015 and 2018 may substantially affect the gain scores displayed in the charts. For example, the Smarter Balanced submission for federal peer review covering spring 2015 tests “revealed some gaps in item coverage at the low end of the performance spectrum.” In January-February 2018, Smarter Balanced released information that the operational item bank used for the spring 2017 testing cycle changed considerably, in an attempt to add easier items to improve coverage at the lower end of the achievement spectrum. Based on Smarter Balanced internal technical data dated October 2016 but not released until January-February 2018, it appeared this effort was not entirely successful. In late March 2018, Smarter Balanced released additional technical information based on analysis of actual spring 2017 item performance. However, this technical information was not consistent with the October 2016 technical information upon which anticipated 2017 item bank performance was based, and the March 2018 technical information did not thoroughly address the differential 2017 Smarter Balanced consortium-wide scores in a way that explained the extremely meaningful declines in scores for 2017. The March 2018 information released by Smarter Balanced as well as reviews of that information by the author are available upon request.

Smarter Balanced has not released information to date on whether the 2018 adaptive item bank and/or adaptive algorithm changed for the 2018 test administration. If there were changes to either the adaptive item bank or adaptive algorithm, that information would inform the interpretation of the data on page 6 considerably. In the absence of such information, one can only conclude that the comparability of the Smarter Balanced testing system from year-to-year remains suspect in terms of generating quality year-to-year change data, with extreme declines in gain scores for 2017 and partial recovery for 2018, substantially different than the PARCC year-to-year gains and at odds with expected annual gains for statewide testing systems.

Finally, it should be noted that the overall consortium testing gain scores for 2016 thru 2018 are lower than the typical gains described by Bob Linn 15 years ago. Over these three years, the Smarter Balanced gains average less than 1.0 percentage point (0.95 to be exact), while the PARCC gains averaged 1.67 percentage points. The letter grades reflect a reasonable mix of A’s and B’s for 2016, but for 2017 and 2018 only a single state each year received a letter grade of B and no state received an A. Why this is the case is unknown. It may be due to the characteristics

of each testing system, or it may be due to the demanding nature of implementing challenging academic content standards in our schools over the past five years.

Author Tagline

Doug McRae is a retired educational measurement specialist living in Monterey, California. In his almost 50 years in the K-12 testing field, he has served as an educational testing company executive in charge of the design and development of K-12 tests widely used across the country, as well as an advisor for the design and development of California's STAR statewide testing system which was used from 1998 through 2013. He has a Ph.D. in Quantitative Psychology from the L. L. Thurstone Psychometric Laboratory at the University of North Carolina, Chapel Hill.